Structure and Stability of Cyclic Peptide Based Nanotubes: An Molecular Dynamics Study of the Influence of Amino Acid Composition

Ramadoss Vijayaraj,[†‡] Sofie Van Damme,[‡] Patrick Bultinck,[‡] and Venkatesan Subramanian[†]*

Chemical Laboratory, CSIR-Central Leather Research Institute, Adyar, Chennai 600 020, India and Department of Inorganic and Physical Chemistry, Ghent University, Krijgslaan 281(S3), Gent 9000, Belgium

Abstract

The stability of self-assembling cyclic peptides (CPs) is attained by the intermolecular backbone-backbone hydrogen bonding (H-bonding) interactions. In addition to this H-bonding interaction, the self-assembled CPs are further stabilized by various intermolecular side chain-side chain interactions. This study investigates the role of amino acids on the structure and stability of self-assembled CPs using classical molecular dynamics (MD) simulations and molecular mechanics/Poisson-Boltzmann surface area (MM/PBSA) method. The amino acids considered for the construction of model structures of cyclic peptide nanotubes (CPNTs) are Ala, Leu, Phe, Gln, Glu, and Trp. The calculated structural parameters from classical MD simulation reveal that the backbone flexibility of CPNTs composed of non-Ala residues is an intrinsic property of the amino acids. The presence of an Ala residue at the alternate position increases the solvation of side chains of Gln residue. The occurrence of Glu residue does not favour the formation of intermolecular side chain-side chain H-bonding interaction in aqueous medium. It is evident from the calculated free energy of binding that the CPNTs composed of non-polar residues are highly stable in aqueous medium. At the same time, the CPNTs with polar side chains are less stable in aqueous medium. Results obtained from this study demonstrate the role played by amino acid side chains on the structure and stability of CPNTs and provide valuable suggestions on the design of CPNTs with moderate stability in various solvent environments.

Keywords: Cyclic peptide, Self-assembly, Hydrogen bond, Molecular dynamics, MM/PBSA

*To whom correspondence should be addressed. E-mail: subuchem@hotmail.com
[†]CSIR-Central Leather Research Institute.
[‡]Ghent University.

Introduction

Cyclic peptides (CPs) with alternate L and D α-amino acids can self-assemble, under appropriate conditions, into hollow, tubular structures by means of ordered antiparallel hydrogen bonding (H-bonding) between the backbones of subsequent chains.[1-4] The structural characteristics of CPs in a cyclic peptide nanotube (CPNT) have been investigated using various experimental techniques and theoretical methods.[1,4,5-16] The backbone carbonyl-amide functionality of CP aligns with the tube axis, thereby stabilizing the self-assembly through H-bonding interactions. The stacking of CPs with opposite orientation, i.e. the CP with N→C orientation stacks with C→N oriented CP, promotes antiparallel β–sheet like H-bonding between unique chiral amino acids. Due to the alternate L- and D- arrangements of the amino acids, the lumen of the CPNT is free from all amino acid side chains. This allows the transport of various ions or small molecules through the lumen of the tube.[3,17-23] The internal diameter of the nanotube can be tailored by changing the number of amino acids in each CP unit. A recent study on such systems shows that the chemical modification of the aromatic residue controls the diameter of the peptide nanotube.[24] Hence, it can be used as a selective transporter of ions and drug molecules with different sizes.[25] The plane formed by the Cα atoms in a CP is called the Cα plane. The region in between two Cα planes is referred to as mid-Cα region. The water molecule arrangement in CPNT formed by octa-cyclic-peptides is identified as one and two in the backbone Cα plane and mid-Cα plane regions, respectively.[7] Recently, it has been demonstrated that UV photoluminescence can be used to determine the number of water molecules in the dipeptide based nanotubes.[26]

The amino acid side chains protrude from the tube surface which facilitates the side chain-side chain interactions. The intermolecular backbone H-bonding interaction provides the primary stabilization to the self-assembled CPs regardless of the presence of the precise

amino acids. The additional stability attained by the intermolecular side chains interactions are modulated by the amino acids involved and their complementary interaction with the solvent molecules. In the presence of bulky side chains, the backbone carbonyl-amide groups of CP are less exposed to the solvent medium from the tube surface.[27] The self-assembled CPNTs containing Gln amino acids with inter- and intra-tubular H-bonding interactions are shown to retain the stability in most common organic solvents and survive even in boiling water.[6]

In the design of CPNTs based on non-standard amino acids, the formation of a tubular structure is found to be compatible with different backbone *N*-alkyl substituents.[28] Further, the calculated association constants for the dimers of D,L-octapeptide reveal that the stability of self-assembled CPs is influenced by the amino acid composition.[28] As the side chain interactions between consecutive chains play an essential role in the stabilization of CPNTs, the overall stability of CPNTs can be engineered by the choice of different amino acids. In the design of CPs, the selection of the amino acid composition is occasionally based on the feasibility of experimental synthesis or to make strong polar/non-polar side chain-side chain interactions.[29] However, extensive studies on the propensity of various amino acids on the structure and stability of CPNTs are scarce. In our previous classical molecular dynamics study, we have investigated the critical steps involved in the self-assembly of CPs based on $\{cyclo[(\text{D-Ala-L-Ala})_4]\}_n$, where n represents the number of CPs ranging from 1 to 8.[27]

In the design of CPNTs based on naturally occurring amino acids (including the D isomer), the Pro and Gly amino acids are least favoured due to inappropriate stacking interaction. The van der Waals and electrostatic interactions between the side chains are the predominant factors which further enhance the stability in addition to the strong backbone-backbone electrostatic interaction. The present study explores various structural properties

and the stability of CPNTs constructed from different amino acids using classical molecular dynamics (MD) simulations, with special attention to the role of the Ala residue. The free energy of binding of various CPNTs has been calculated by employing a molecular mechanics/Poisson-Boltzmann surface area (MM/PBSA) approach. The composition of amino acids used for the design of CPs by various experimental and theoretical studies is given in Table 1. Based on the reported theoretical and experimental studies on CPNTs, the amino acids considered for the construction of CPNT are Ala, Leu, Phe, Gln, Glu, and Trp. Though a CP comprising eight amino acids can have 8^{18} possible amino acid sequences excluding Pro and Gly amino acids, the side chain clash between the adjacent amino acids reduces the number of possible combinations. The present study investigates the structure and stability of CPNTs previously studied using experimental and/or theoretical methods. The obtained results provide valuable suggestions for the design of energetically more stable CPNTs for various applications.

Computational details

The composition of various CPNTs considered for the present study and the corresponding nomenclature are presented in Table 2. All model systems were constructed by self-assembling eight CPs with antiparallel stacking. The AF, AL, AQ, QL, and WL model systems exhibit C_4 symmetry. The QAEA and WL3QL model systems exhibit C_2 and C_1 symmetries, respectively. The side chain of the Glu residue in the QAEA system was protonated to avoid the repulsive electrostatic interaction between the negatively charged carboxylate groups. Prior to the MD simulation, all initial conformations were subjected to energy minimization in an implicit solvent environment to reduce unfavourable short contacts. Each CPNT was solvated with a cubic box of TIP3P water molecules extending 16 Å away from the solute atoms.[37,38] The energy minimization, equilibration and production

simulations were carried out in different stages: (a) all atom minimization for 10000 cycles, (b) equilibration of the solvent for 200 ps under position restraints of the solute atoms through a harmonic force constant of 1000 kJ nm^{-2} in the NVT ensemble followed by further equilibration in the NPT ensemble in the absence of any positional restraints, (c) production MD simulation for 10 ns in the NPT ensemble using a 2 fs time step. From the production run, the structural information was collected every 0.2 fs. All simulations were carried out with periodic boundary conditions applied in three dimensions. The pressure was controlled at 1 atm using a Parrinello-Rahman barostat and the temperature of the system was maintained at 300 K with a velocity rescaling thermostat. The protein and non-protein atoms were coupled to separate temperature coupling baths. The particle-mesh Ewald (PME) summation method was used for calculating the long-range electrostatic interactions.[39] The short-range and long-range non-bonded interactions were truncated at 1.4 and 12 Å, respectively. The linear constraint solver (LINCS) algorithm was used to constrain the bonds involving hydrogen atoms.[40] All the calculations were carried out using GROMACS 4.5.1[41,42] employing ff99SB[43] force field parameters.

MM/PBSA calculations

The free energy of binding of each CP unit in a CPNT with the remaining CP units as described in our previous study and the total free energy of binding was calculated using the MM/PBSA methodology as implemented in AmberTools 1.5.[27,44,45] The free energy of binding was averaged from 200 snapshots extracted from the last 2 ns production simulation with a 10 ps step size. The explicit water molecules were removed from each snapshot, and the free energy of binding of each CP unit with other CP units of the same system was calculated using

$$\Delta G^{n}_{binding} = G_{complex} - (G^{n} + G^{m}_{complex}) \qquad (1)$$

where G^{n} represents the energy of the nth monomer in the various CPNTs and the position of the n ranges from CP1 to CP8 (Figure 1). $G^{m}_{complex}$ denotes the energy of the complex which excludes the nth monomer (G^{n}) and $G_{complex}$ represents the energy of the complex. The $\Delta G^{n}_{binding}$ is the binding free energy for the nth monomer CP unit. The free energy of binding of each subunit was calculated using eq. 2.

$$\Delta G_{binding} = \Delta E_{MM} + \Delta G_{solv} - T\Delta S_{solute} \qquad (2)$$

The molecular mechanics energy, ΔE_{MM} was divided into

$$\Delta E_{MM} = \Delta E_{internal} + \Delta E_{vdW} + \Delta E_{ele} \qquad (3)$$

Where $\Delta E_{internal,}$ ΔE_{vdW} and ΔE_{ele} represents internal, van der Waals and the electrostatic contributions to the MM energy. The solvation energy was divided into two terms as shown in eq. 4.

$$\Delta G_{solv} = \Delta G_{pol} + \Delta G_{np} \qquad (4)$$

The polar contribution (ΔG_{pol}) to the free energy of solvation (ΔG_{solv}) was calculated by solving the Poisson-Boltzmann (PB) equation. The nonpolar contribution (ΔG_{np}) to the solvation free energy was calculated from the solvent accessible surface area (SASA) using the linear combination of pairwise overlaps (LCPO) method[46] as implemented in Sander program, according to eq. 5.

$$\Delta G_{np} = \gamma SASA + \beta \qquad (5)$$

where the surface tension (γ) and the offset (β) were set to the default values of 0.00542 kcal mol^{-1} $Å^{-2}$ and -1.008 kcal/mol, respectively. The interior and exterior dielectric constants were set to 1 and 80, respectively, with a grid spacing set to 0.5 Å and 1000 linear iterations were performed. Bond radii and a probe radius of 1.4 Å were used for both ΔG_{pol} and ΔG_{np} calculations. The harmonic approximation of the translational, rotational and vibrational conformational entropies to the free energy of binding was calculated using the normal-mode analysis program of MM/PBSA package. The normal-mode analysis was performed on 10 snapshots extracted from the last 2 ns production simulation with 200 ps step size. In addition to the calculation of the free energy of binding of each CP unit in various CPNTs, the total free energy of binding of each CPNT was calculated using eq. 6.

$$\Delta G_{binding}^{tot} = G_{complex} - \left(\sum_{CPn} G_{monomer} \right) \qquad (6)$$

where $G_{complex}$ represents the energy of the complex. The $G_{monomer}$ denotes the CP monomer energy where n ranges from 1 to 8. The $G_{complex}$ energy in eq. 1 is invariant for each CPNT and $G_{complex}$ and $G_{monomer}$ energies were obtained from $G_{complex}$ and G^n, respectively of eq. 1. The $\Delta G_{binding}^{tot}$ represents the total free energy of binding of various CPNTs.

Results and discussion

Dynamical Properties of CPNTs

The stability of CPNTs is determined by the intermolecular backbone-backbone H-bonding and various side chain-side chain interactions. The presence of the different amino acids and their associated side chain-side chain interactions between two CPs influence the dynamic nature of the backbone C=O···H–N interaction. Therefore, it is necessary to explore

the effect of the side chain on the structural stability of CPNTs. In this context, the RMSD values of the heavy atoms of the backbone and the side chain atoms were calculated with respect to the initial geometry from the MD trajectory. The calculated RMSD values from the present study are plotted in Figure 2. The average RMSD of the backbone atoms and its corresponding standard deviation shows that the backbone fluctuations depend on the nature of the side chains. Further, the calculated RMSDs of various CPNTs containing Ala in the alternate position are almost similar to that of the AA system. However, in the absence of the Ala residue (QL, WL and WL3QL), the backbone fluctuations are an intrinsic property of the amino acid. Thus, RMSD analysis reveals that the combination of various residues would provide different structural stability to the CPNTs. It can be noted that the QL system has the lowest RMSD value corresponding to the backbone atoms when compared to other systems.

The calculated RMSD for the different side chains shows that AA has the lowest value and AF exhibits the highest value. In CPNTs comprising Ala and other residues, the presence of Ala with other residues increases the flexibility of CPNTs. For instance, the side chain dihedral angle (χ^1) distribution of the Phe residues in AF system has been measured from the final snapshot of MD simulation. The measured χ^1 values are -170±10º or 170±10º. Hence, the AF system exhibits the highest deviation in the RMSD values. It is evident from the RMSD values of the side chain of the AQ, QAEA and QL systems that the presence of Leu residue at the alternate positions imposes dynamical restrictions and thereby the flexibility of QL decreases when compared to the AQ and QAEA systems. The comparison of RMSD of backbone and side chain atoms reveals the following: (i) the ordered H-bonding interaction between the backbone atoms of any two consecutive CPs reduces the backbone flexibility and (ii) since the side chain has no structural restrictions (lies outside the tube), it tries to acquire optimal positions for the formation of stable CPNT. Thus the RMSD of side

chain atoms increases. As mentioned earlier, the stability of CPNTs depends predominantly on strong backbone-backbone H-bond interactions. However, the interactions between side chains can also affect the stability of nanotubes. For example, a Gln or Glu residue in CPNT can form hydrogen bonds not only with the backbone atoms but also between side chains. These hydrogen bonds between side chains enhance the stability of CPNTs. Thus the fluctuations in these side chains may decrease upon formation of CPNTs.

Structural Features

The average distance between the centre of mass of the first and last Cα planes is defined as the length of the CPNT. The calculated average tube lengths along with their standard deviations for CPNTs considered in the present study are presented in Figure 3. The AA system forms a tube of shortest length (33.37 Å). Earlier electron diffraction and MD studies reported a intermolecular distance of 4.73–4.80 Å for the self-assembled CPs composed of different amino acids.[1,4,6,27,30] Extrapolating this value gives 33.11–33.60 Å as tube length for a system consisting of eight CPs which is very close to the present result. We have taken the average tube length of the AA system as a reference for other CPNTs due to the minimal perturbation on the backbone-backbone H-bonding induced by the Ala residues. In addition, it has least intermolecular side chain-side chain interaction due to the presence of only a relatively small side chain (-CH_3) when compared to the other amino acids considered in this study. It can be seen that the tube length of the AL, AQ and QAEA systems is comparable to that of AA. On the other hand, the presence of the Trp residue increases the tube length by ~ 0.6 Å when compared to AA. The tube length of the AF and QL CPNTs is marginally higher than that of AA by 0.22 and 0.25 Å, respectively. These results show that the side chain-side chain interactions play a role in determining the length of tube. Overall, in CPNTs, the tube length is directly proportional to the backbone-backbone H-bonding

interaction. In addition, the presence of bulky side chains prevents the close packing of CPs in CPNTs which in turn marginally increases the length of the tube.

In order to explore the role of the side chains in the self-assembly, the average distance between the Cα planes of each consecutive CP was measured (See Table S1 of Supporting Information). All the inter subunit distances in the AA system are nearly equal to 4.78 Å in the inner region (CP2 to CP7). In case of the WL and WL3QL systems, the intermolecular Cα plane distance marginally varies from one plane to other. It is interesting to note an odd-and even variation in the Cα plane distances of AF, AL, AQ, QAEA and QL systems. This can be directly related to the interaction between the side chains of the successive CP units. The presence of amino acid residues with bulky side chains is responsible for the observed odd-even effect. It can be seen from Figure 4 that the intermolecular stacking in CPNT is characterized by alternating backbone-backbone and side chain-side chain interactions. The occurrence of Ala or Leu residues at the alternate positions minimizes the side chain hindrance and in turn reduces intermolecular stacking distance.

The diameter of the CP was calculated as the longest distance between any two Cα atoms. The average diameters and their standard deviations for various model systems are presented in Table S2 of Supporting Information. The calculated average diameter for the AA system is 9.77 Å which is comparable with the results of a previous MD study.[27] The increase in the average diameters of the other CPNTs ranges from 0.2 to 0.7 Å. It can be noted from these results that the incorporation of other amino acids slightly increases the average diameter of the CPNT when compared to the AA system.

The relative orientation of the backbone atoms in each CP unit is measured using the dihedral angle formed by the alternate Cα atoms in each CP unit. Two such dihedral angles can be measured for a CP containing eight amino acids. A dihedral angle equal to 0º indicates

that all Cα atoms are in the same plane. The deviations in the calculated dihedral angle from 0° reflect the twist in the CPs. The probability of finding a particular dihedral angle in each CP unit for various model systems is presented in Figure 5. It is clear from Figure 5 that the dihedral angle at the core regions of different CPNTs varies from -7 to 7° except for the WL system. In case of the AA, AQ, AF and QAEA systems, the maximum probability is observed for 0°. For other CPNTs, the maximum of distribution ranges from -4 to 4°.

Intermolecular backbone H-bonds

A CP unit consisting of eight amino acids has eight H-bond donor and acceptor groups in the backbone. Excluding the free carbonyl and amide groups in the terminal of CPs, CPNTs consisting of eight CP units contain 56 unique intermolecular backbone H-bonds. These intermolecular backbone H-bonds are often perturbed by two factors: (i) the interaction of water molecules with the carbonyl-amide functionality and (ii) the occurrence of bulky side chains increases the intermolecular distance due to intermolecular side chains hindrance. MD studies on similar systems have demonstrated that the presence of small side chains such as Ala entails accessibility of water molecules to the backbone carbonyl group from the surface of the tube.[27] The normalized frequency of occurrence of all intermolecular backbone H-bonds in each frame of last the 5 ns of the MD trajectory was analysed for various model systems. Here, the original data was normalized to 15. A H-bond distance ≤ 3.5 Å and a H-bond angle ≤ 150° was used as criterion to identify the existence of H-bonds. The results are shown in Figure 6. It is evident from Figure 6 that the AA system has the highest frequency with respect to the H-bond range corresponding to 48-54. At the same time, the frequency of simultaneous formation of 55 and 56 H-bonds in AA system is less than that of other systems. The QL system forms most H-bonds compared to all other systems. The frequencies of all other systems range between the QL and AA systems. The factors responsible for the

reduction in the formation of number of backbone-backbone H-bonded interaction are: (i) the H-bond donor and acceptors groups are not optimally placed due to the side chain-side chain clash, and (b) the possibility of interaction of water molecules with the backbone carbonyl groups.

Interaction of CPNT with water molecules

The analysis of H-bonds between the backbone carbonyl-amide functionalities with water molecules enhance the understanding of the H-bonding interaction in various CPNTs. Figure 7 shows the normalized frequency of H-bonds observed between the carbonyl O atom and water molecules from the last 1ns MD simulation of different model systems. It is evident from Figure 7 that the backbone carbonyl O atoms of the AA system form the highest number of H-bonds with water molecules compared to other systems. MD studies on CPNT consisting of eight CP units have reported that the Cα and mid-Cα planes can accommodate one and two water molecules, respectively.[30,32,34] Thus the octamer CPNT can accommodate ~22 water molecules in the lumen of the tube. The frequency of occurrence of H-bonds presented in Figure 7, also includes the carbonyl O atoms in the terminal CP that do not form intermolecular H-bonding interaction with other CPs. The H-bonds between the carbonyl O atom and water molecules in all model systems can be classified into three categories based on the frequency of occurrence of H-bonding interaction: (i) maximum, (ii) intermediate, and (iii) minimum. It can be seen that the carbonyl group in AA forms the highest number of H-bonds with the water molecules. About 35-41 H-bonds are formed in the AF, AL, AQ and QAEA systems. The QL, WL and WL3QL systems form marginally fewer H-bonds with the water molecules due to the presence of a bulky side chain that hampers the interaction between the carbonyl group and water molecule from the surface of the tube. Therefore the

amino acid composition of CPs plays a part in the formation of backbone-backbone H-bonding as well as backbone carbonyl-water H-bonding interaction.

Another factor that can influence the intermolecular backbone H-bonding interaction is the interaction of water molecules with the backbone amide functionality (N-H) from the surface of the tube. The normalized frequency of occurrence of these types of H-bonds is given in Figure 8. It is evident from Figure 8 that the maximum frequency for backbone amide-water H-bonding (N–H···O–H) varies from 5 to 7 for various CPNTs. Therefore the amide-water interaction plays only a marginal role in the disruption of backbone-backbone H-bonding interaction. Previous studies on similar systems demonstrated that there is no specific interaction between the backbone amide groups and water molecules.[30]

Intermolecular side chains H-bond

The presence of polar side chains facilitates the formation of inter CP side chain H-bonding interaction. This H-bonding interaction enhances the stability of the CPNT in addition to the primary backbone-backbone H-bonding interaction. Therefore, the number of H-bonds formed between the side chains of two Gln residues in the octamer on the one hand and the side chain of Gln with water molecules on the other hand were calculated from the MD trajectory. The results obtained are depicted in Figures 9 and 10. The frequency of occurrence of a H-bonding interaction between the side chains of one Gln with another Gln is less than that of its interaction with water molecules. In the case of the AQ system, the number of intermolecular H-bonds between the Gln side chains ranges from 0-12 (Figure 9) while the number of Gln-water H-bonds varies from 86 to 130 (Figure 10).

Even though the AQ and QL systems contain the same number of Gln residues, the QL system shows a higher preference for intermolecular Gln-Gln side chain H-bonding

interaction compared to the AQ system. At the same time, the Gln-water H-bonding interaction in the QL system is lower than in the AQ system. The observed differences in these two systems may be attributed to the presence of a Leu residue that restricts the access of water molecules to the side chains of the Gln residue which in turn facilitates the formation of the H-bonding interaction between the two Gln residues.

The QAEA and WL3QL systems have a small number of intermolecular Gln-Gln H-bonding interactions than the AQ and QL systems in agreement with their amino acid composition. The occurrence of Ala at alternate positions of QAEA facilitates the access of water molecules to the Gln amino acids. Thus the interaction of water molecules with QAEA system is significantly higher than the WL3QL system (Figure 10). There are no intermolecular H-bonds observed between the side chains of the Glu residues in the QAEA system. This shows that the protonation of the carboxylate groups does not favour the intermolecular side chain-side chain H-bonding interaction in aqueous medium. However, the Glu side chains are H-bonded with water molecules (Figure 10). It is evident from Figure 10 that the N–H group at the indole ring of Trp interacts through H-bonding with water molecules.

Free energy of binding

The different energy contributions to the free energy of binding of each CP unit are presented in Table S3 of Supporting Information. The binding energy of each CP using the PB method and free energy of binding of the model systems are depicted in Tables 3 and 4, respectively. In case of the AA system, the calculated energy components agree with the previous MD study.[27] It is evident from the results that the van der Waals contribution (ΔE_{vdW}) of each CP unit varies with the presence of different hydrophobic side chains. The AA system has the lowest van der Waals contribution in comparison to all other CPNTs. The

van der Waals contributions of the AL, AQ and QAEA systems are almost similar. The presence of the Trp residue substantially increases the van der Waals contribution of the WL and WL3QL systems. Since the electrostatic interactions in the AA, AF, AL and WL systems are predominantly based on the backbone-backbone H-bonding interactions, the electrostatic contributions in these systems are similar. Due to the possibility of side chain-side chain H-bonding interactions in other systems (AQ, QAEA, QL, and WL3QL), the electrostatic contribution varies from one system to other. The standard deviation in the electrostatic contribution of AQ, QAEA, QL, and WL3QL is comparatively higher than for the other models systems. The presence of polar side chains increases the polar contribution (ΔG_{pol}) to the solvation free energy. The solvation contributions to the total van der Waals ($\Delta G_{vdW}^{tot} = \Delta E_{vdW}^{tot} + \Delta G_{np}^{tot}$) and electrostatic ($\Delta G_{ele}^{tot} = \Delta E_{ele}^{tot} + \Delta G_{pol}^{tot}$) terms are given in Table 5. The van der Waals and non-polar solvation contributions (ΔG_{vdW}^{tot}) favour the self-assembling process of CPs and these contributions are directly proportional to the size of the non-polar side chain.

Results from the PB method (ΔG_{PB}) reveal that the binding energy of each CP unit at the core regions of the AA system is quite similar (Table 3). The WL and WL3QL systems have the highest binding energy. This is due to the contribution from the intermolecular side chains van der Waals interaction in addition to the backbone-backbone H-bonding interaction. It can be noted that the standard deviations of all systems are slightly higher than that of the AA system which is in agreement with the orientation of the side chain and its interaction with subsequent units. Even though the QL system has higher occurrence of intermolecular H-bonds compared to all other systems, the van der Waals interaction contribution stabilizes WL and WL3QL systems when compared to all other systems. It is clear from the binding energies of the CPNTs that the hydrophobic side chains increase the

stabilization energy in the polar environment. Although the H-bonding between the side chains of various CPNTs favours the self-assembly process, the solvation effect reduces the contribution of side chain H-bonding to the free energy of binding. The calculated entropies of the systems reveal that conformational entropic contributions arising from the core regions are higher than those of the terminal regions. This may be attributed to the restrictions in the rotational and translational motions of the side chains (Table S3). The entropy contribution decreases with respect to the increase in the size of the side chains (Table 5). The free energy of binding ($\Delta G_{binding}$) of each CP unit at the core region of the AA system, as depicted in Table 5, ranges from -34 to -37 kcal/mol. These values are very close to the previous molecular dynamics results.[27]

The free energy of binding at the terminal CPs can be considered as the effective measure of the self-assembling process. The increase in the free energy of binding in the terminal CP unit enhances the rate of self-assembling process. The analysis of free energy of binding at terminal CPs reveals that the presence of the Trp residue increases the free energy of terminal CPs (~-24 kcal/mol) which is substantially higher than that of the AA system (Table 4). The decomposition of the free energy of binding with respect to each CP unit shows that the free energy of binding of one CP unit to another in the core region is different for the addition of another CP, except in the case of AA. Therefore, the amino acid composition determines the stability of the CPNTs. The total free energy of binding ($\Delta G_{binding}^{tot}$) presented in Table 5 highlights the propensity of the different residues to form CPNTs and their combinations.

The AA system has the lowest free energy of binding and the WL system has the highest free energy of binding. On account of negligible intermolecular side chains interaction, the AA system exhibits the lowest stability when compared to the other model

systems. In aqueous medium, the intermolecular electrostatic interaction is opposed by the solvation effect. At the same time, the intermolecular hydrophobic interaction is favoured by the presence of non-polar amino acids. The QL, WL and WL3QL systems have twofold higher free energy of binding than that of the AA system due to various intermolecular side chains interactions. It is evident from the results that the presence of different amino acids and concomitant intermolecular side chain interactions significantly improve the stability of CPNTs.

Conclusion

The present study characterizes the propensity of amino acids to form stable CPNTs using classical MD simulation and calculation of free energy binding using the MM/PBSA method. The RMSD of the backbone atoms shows a similar trend for the CPs containing an Ala residue in every (alternate) position. In the absence of an Ala residue, the backbone flexibility is an intrinsic property of amino acids. There is no specific correlation observed between the flexibility of the backbone and the side chain atoms. The length of the CPNT depends on the bulky nature of the side chains in addition to the backbone-backbone H-bond interaction. The AA model system forms the tube with shortest length. The length of the WL and WL3QL systems are ~1 Å longer than that of the AA system. The presence of the Ala amino acid facilitates the access of water molecules to the backbone carbonyl-amide groups from the surface of the tube. Thereby the CPNT with an Ala residue at every (alternate) position has less free energy of binding when compared to that of the QL, WL, and WL3QL model systems. In addition, the Ala residue at the alternate positions increases the solvation of side chain of Gln residues. Hence, the tendency for intermolecular side chain-side chain H-bonding involving Gln residues is reduced in the presence of Ala. Non-polar amino acids are found to increase the van der Waals energy contribution to the free energy of binding. Amino

acids with polar side chain augment the solvation of CPNT in aqueous medium rather than the intermolecular H-bonding interaction. On the basis of free energy of binding, the stability of various CPNTs in aqueous medium can be ordered as AA < QAEA < AQ < AL < AF < QL < WL3QL < WL.

Acknowledgements.

We thank the Council of Scientific and Industrial Research (CSIR), New Delhi, for financial assistance. R.V. is grateful to the Erasmus Mundus External Cooperation Window, Eurindia project for the financial assistance. R. V. and V. S. Acknowledge the Multi-Scale Simulation and Modeling project (MSM) for providing financial assistance. The computational resources (STEVIN Supercomputer Infrastructure) and services used in this work were kindly provided by Ghent University, the Flemish Supercomputer Center (VSC), the Hercules Foundation and the Flemish Government – department EWI.

Electronic Supplementary Information:

Average distance between the centre of mass of Cα planes of consecutive chains in various CPNTs. Average diameter of CPs. Energy contributions to the free energy of binding.

References

1. M. R. Ghadiri, J. R. Granja, R. A. Milligan, D. E. McRee, N. Khazanovich, *Nature* 1993, **366,** 324-327.
2. M. R. Ghadiri, J. R. Granja, L. K. Buehler, *Nature* 1994, **369,** 301-304.
3. J. R. Granja, M. R. Ghadiri, *J. Am. Chem. Soc.* 1994, **116,** 10785-10786.
4. N. Khazanovich, J. R. Granja, D. E. McRee, R. A. Milligan, M. R. Ghadiri, *J. Am. Chem. Soc.* 1994*,* **116,** 6011-6012.
5. M. R. Ghadiri, K. Kobayashi, J. R. Granja, R. K. Chadha, D. E. McRee, *Angew. Chem. Int. Ed. Engl.* 1995, **34,** 93-95.
6. J. D. Hartgerink, J. R. Granja, R. A. Milligan, M. R. Ghadiri, *J. Am. Chem. Soc.* 1996, **118,** 43-50.
7. C. Gailer, M. Feigel, *J. Comput.-Aided Mol. Des.* 1997, **11,** 273-277.
8. J. P. Lewis, N. H. Pawley, O. F. Sankey, *J. Phys. Chem. B* 1997, **101,** 10576-10583.
9. R. A. Jishi, R. M. Flores, M. Valderrama, L. Lou, J. Bragin, *J. Phys. Chem. A* 1998, **102,** 9858-9862.
10. J. D. Hartgerink, T. D. Clark, M. R. Ghadiri, *Chem. Eur. J.* 1998, **4,** 1367- 1372.
11. F. Ferrante, G. L. Manna, *J. Mol. Struc-Theochem* 2003, **634,** 181–186.
12. H. Tan, W. Qu, G. Chen, *Int. J. Quantum Chem.* 2005, **102,** 1106-1115.
13. J. Zhu, J. Cheng, Z. Liao, Z. Lai, B. Liu, *J. Comput. Aided Mol. Des.* 2008, **22,** 773–781.
14. J. Cheng, J. Zhu, B. Liu, Z. Liao, Z. Lai, *Mol. Simulat.* 2009, **35,** 625–630.
15. W. Qu, H. Tan, G. Chen, R. Liu, *Int. J. Quantum Chem.* 2010, **110,** 1648–1659.
16. R. Hourani, C. Zhang, R.V.D. Weegen, R. Ruiz, C. Li, S. Keten, B.A. Helms, T. Xu, *J. Am. Chem. Soc.*2011, **133,** 15296–15299.
17. H. S. Kim, D. Hartgerink, M. R. Ghadiri, *J. Am. Chem. Soc.* 1998, **120,** 4417–4424.
18. T. D. Clark, L. K. Buehler, M. R. Ghadiri, *J. Am. Chem. Soc.* 1998, **120,** 651-656.
19. J. Sanchez-Quesada, M. P. Isler, M. R. Ghadiri, *J. Am. Chem. Soc.* 2002, **124,** 10004-10005.
20. D. Asthagiri, D. Bashford, *Biophys. J.* 2002, **82,** 1176–1189.
21. H. Hwang, G. C. Schatz, M. A. Ratner, *J. Phys. Chem. B* 2006, **110,** 6999-7008.
22. H. Hwang, G. C. Schatz, M. A. Ratner, *J. Phys. Chem. B* 2006, **110,** 26448-26460.
23. F. Dehez, M. Tarek, C. Chipot, *J. Phys. Chem. B Lett.* 2007, **111,** 10633-10635.
24. C. Tarabout, S. Roux, F. Gobeaux, N. Fay, E. Pouget, C. Meriadec, M. Ligeti, D. Thomas, M. IJsselstijn, F. Besselievre, D. A. Buisson, J. M. Verbavatz, M. Petitjean,

C. Valéry, L. Perrin, B. Rousseau, F. Artzner, M. Paternostre, J. C. Cintrat, *Proc. Natl. Acad. Sci. USA,* 2011, **10,** 7679-7684.

25. F. Rahmat, N. Thamwattana, B. J. Cox, *Nanotechnology* 2011, **22**, 1-8.
26. M. Wang, S. Xiong , X. Wu , P. K. Chu, *small* 2011, **7,** 2801–2807.
27. R. Vijayaraj, S. S. Raman, R. M. Kumar, V. Subramanian, *J. Phys. Chem. B* 2010, **114,** 16574-16583.
28. T. D. Clark, J. M. Buriak, K. Kobayashi, M. P. Isler, D. E. McRee, M. R. Ghadiri, *J. Am. Chem. Soc*. 1998, **120,** 8949-8962.
29. H. Liu, J. Chen, Q. Shen, W. Fu, W. Wu, *Mol. Pharmaceutics* 2010, **7,** 1985-1994.
30. M. Engels, D. Bashford, M. R. Ghadiri, *J. Am. Chem.* Soc. 1995, **117,** 9151-9158.
31. G. Chen, S. Su, R. Liu, *J. Phys. Chem. B* 2002, **106,** 1570-1575.
32. M. Tarek, B. Maigret, C. Chipot, *Biophys. J.* 2003, **85,** 2287-2298.
33. E. Khurana, S. O. Nielsen, B. Ensing, M. L. Klein, *J. Phys. Chem. B* 2006, **110,** 18965-18972.
34. J. Liu, J. Fan, M. Tang, W. Zhou, *J. Phys. Chem. A* 2010, **114,** 2376–2383.
35. J. Liu, J. Fan, M. Tang, M. Cen, J. Yan, Z. Liu, W. Zhou, *J. Phys. Chem. B* 2010, **114,** 12183–12192.
36. J. Liu, J. Fan, M. Cen, X. Song, D. Liu, W. Zhou, Z. Liu, J. Yan *J. Chem. Inf. Model.* 2012, DOI: 10.1021/ci300185c.
37. W. L. Jorgensen, *J. Am. Chem. Soc.* 1981, **103,** 335–340.
38. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* 1983, **79,** 926–935.
39. T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* 1995, **103,** 8577–8593.
40. B. Hess, H. Bekker, H. J. C. Bendersen, J. G. E. M. Fraaije, *J. Comput. Chem.* 1997, **18,** 1463–1472.
41. B. Hess, C. Kutzner, D. V. D. Spoel, E. Lindahl, *J. Chem. Theory Comput.* 2008, **4,** 435–447.
42. http://www.gromacs.org/ (Last accessed on 13 June 2012)
43. V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, *Proteins* 2006, **65,** 712–725.
44. D. A. Case, T. A. Darden, III T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossvai, K. F. Wong, F. Paesani, J. Vanicek, J. Liu, X.

Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M. J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, P. A. Kollman. (2010), AMBER 11, University of California, San Francisco.

45. http://www.ambermd.org/ (Last accessed on 13 June 2012)

46. J. Weiser, P. S. Shenkin, W. C. Still, *J. Comput. Chem.* 1999, **20,** 217–230.

Table 1: The CP amino acids composition used in previous experimental and theoretical studies.

No.	CP amino acid composition	Reference
1.	*cyclo*[(D-Ala-L-Gly)$_4$]	7,[a] 9[a]
2.	*cyclo*[(D-Ala-L-Ala)$_4$]	14,[a, b] 15,[a], 25[a], 27[a]
3.	*cyclo*[(D-Ala-L-Gln)$_4$]	6,[b] 30[a]
4.	*cyclo*[(D-Ala-L-Leu)$_4$]	6[b]
5.	*cyclo*[(D-Ala-L-Phe)$_4$]	13,[a, b] 15,[a] 31[a]
6.	*cyclo*[(L-Gln-D-Ala-L-Glu-D-Ala)$_2$]	1,[b] 4,[b] 8,[a] 22[a]
7.	*cyclo*[(D-Gln-L-Leu)$_4$]	6[b]
8.	*cyclo*[(D-Trp-L-Leu)$_3$-D-Gln-L-Leu]	26,[a] 32[a]
9.	*cyclo*[(D-Trp-L-Leu)$_4$]	21,[a] 23,[a] 33-36[a]

[a]theoretical study, [b]experimental study

Table 2: The CP amino acid sequence, nomenclature, and composition of the model systems constructed in the present work.

CPs	name	total number of water molecules	total number of atoms
{*cyclo*[(D-Ala-L-Ala)$_4$]}$_8$	AA	10622	32506
{*cyclo*[(D-Ala-L-Phe)$_4$]}$_8$	AF	13505	41475
{*cyclo*[(D-Ala-L-Leu)$_4$]}$_8$	AL	12558	38602
{*cyclo*[(D-Ala-L-Gln)$_4$]}$_8$	AQ	12855	39429
{*cyclo*[(L-Gln-D-Ala-L-Glu-D-Ala)$_2$]}$_8$	QAEA	13521	41411
{*cyclo*[(D-Gln-L-Leu)$_4$]}$_8$	QL	12534	38754
{*cyclo*[(D-Trp-L-Leu)$_4$]}$_8$	WL	15305	47291
{*cyclo*[(D-Trp-L-Leu)$_3$-D-Gln-L-Leu]}$_8$	WL3QL	15125	46695

Table 3: The binding energy (in kcal/mol) of each CP unit as obtained from the PB method for all CPNT systems.

	ΔG_{PB}[a]															
CP unit	AA		AF		AL		AQ		QAEA		QL		WL		WL3QL	
	AVG	SD	AVG	SD	AVG	SD	AVG	SD	AVG	SD	AVG	SD	AVG	SD	AVG	SD
CP1	-27.18	1.98	-39.30	2.58	-34.21	2.51	-34.82	3.06	-34.51	2.60	-45.84	3.75	-50.86	3.20	-49.31	3.42
CP2	-54.63	2.79	-77.73	3.42	-71.08	3.45	-67.59	3.64	-66.08	4.04	-92.83	5.01	-101.38	3.96	-98.61	4.56
CP3	-55.80	2.57	-78.74	3.64	-72.61	3.27	-70.40	4.37	-66.61	3.64	-97.54	5.21	-107.30	4.15	-105.97	4.68
CP4	-55.85	2.77	-74.29	3.39	-71.02	3.76	-69.80	4.03	-68.09	4.04	-96.88	4.98	-106.07	4.22	-105.45	4.00
CP5	-55.57	2.60	-73.75	3.56	-69.70	3.74	-68.70	3.97	-68.95	4.22	-95.24	4.65	-109.30	4.08	-100.30	4.54
CP6	-55.57	2.66	-77.49	3.54	-73.27	3.14	-69.72	4.87	-69.29	3.46	-94.74	4.43	-110.19	3.68	-101.46	4.67
CP7	-54.03	3.01	-75.80	3.66	-71.51	3.48	-66.04	4.43	-65.72	3.37	-92.01	4.20	-102.46	4.26	-100.34	4.08
CP8	-26.48	2.30	-37.92	2.44	-32.65	2.64	-32.56	2.59	-32.12	2.51	-46.15	3.04	-48.87	3.25	-49.61	3.14

[a] ΔG_{PB} is obtained from the addition of ΔE_{MM} and ΔG_{solv}.

Table 4: The free energy of binding (in kcal/mol) of each CP unit for all CPNT systems.

	$\Delta G_{binding}$[a]															
CP unit	AA		AF		AL		AQ		QAEA		QL		WL		WL3QL	
	AVG	SD	AVG	SD	AVG	SD	AVG	SD	AVG	SD	AVG	SD	AVG	SD	AVG	SD
CP1	-10.76	2.39	-18.92	4.82	-15.79	2.82	-16.49	3.27	-15.96	2.94	-22.22	4.06	-24.41	3.53	-23.39	3.94
CP2	-37.45	4.17	-56.38	5.98	-53.33	4.49	-49.28	4.68	-45.91	4.21	-66.78	5.56	-75.03	4.44	-69.21	5.11
CP3	-36.73	4.01	-56.87	4.95	-52.58	3.61	-52.36	4.77	-47.04	4.60	-71.22	5.29	-81.81	4.43	-81.03	4.85
CP4	-37.10	3.16	-52.69	4.98	-49.48	5.48	-51.95	4.27	-47.73	5.65	-71.13	5.23	-77.98	4.96	-77.58	4.52
CP5	-37.83	4.08	-52.25	6.55	-48.92	5.93	-49.68	5.15	-49.94	5.47	-69.62	6.67	-86.44	5.37	-73.09	5.18
CP6	-36.43	4.33	-57.43	6.12	-52.47	4.83	-49.74	5.70	-49.66	4.32	-70.64	5.31	-83.77	4.39	-75.17	5.12
CP7	-34.86	5.89	-55.81	5.09	-51.10	3.95	-45.72	4.72	-45.22	3.79	-67.54	5.81	-77.04	4.80	-73.98	4.18
CP8	-10.10	2.53	-18.35	3.48	-14.45	3.27	-13.78	3.28	-14.04	2.60	-22.93	3.55	-25.34	3.39	-24.62	4.15

[a] $\Delta G_{binding}$ is calculated from ΔG_{PB}- $T\Delta S$.

Table 5: Contribution of the energy components (in kcal/mol) to the total free energy of binding for all CPNTs.

	AA	AF	AL	AQ	QAEA	QL	WL	WL3QL
ΔE_{vdW}^{tot}	-133.57	-224.51	-194.33	-188.14	-189.39	-268.92	-353.51	-336.22
ΔE_{ele}^{tot}	-296.97	-289.56	-289.59	-344.33	-336.32	-393.23	-303.62	-320.20
ΔE_{MM}^{tot}	-430.54	-514.06	-483.92	-532.46	-525.70	-662.15	-657.13	-656.42
ΔG_{pol}^{tot}	253.66	266.14	251.91	313.42	310.44	356.44	313.72	326.18
ΔG_{np}^{tot}	-13.42	-22.07	-18.84	-18.73	-18.18	-25.83	-31.09	-29.78
ΔG_{solv}^{tot}	240.24	244.07	233.07	294.69	292.27	330.61	282.63	296.39
ΔG_{vdW}^{tot}	-146.99	-246.58	-213.17	-206.87	-207.56	-294.75	-384.60	-366.00
ΔG_{ele}^{tot}	-43.31	-23.41	-37.68	-30.90	-25.87	-36.78	10.09	5.98
ΔG_{PB}^{tot}	-190.30	-269.99	-250.85	-237.77	-233.44	-331.54	-374.50	-360.02
$T\Delta S^{tot}$	-113.78	-157.45	-141.63	-137.37	-141.19	-169.93	-183.15	-183.26
$\Delta G_{binding}^{tot}$	-76.52	-112.54	-109.22	-100.40	-92.25	-161.61	-191.35	-176.76

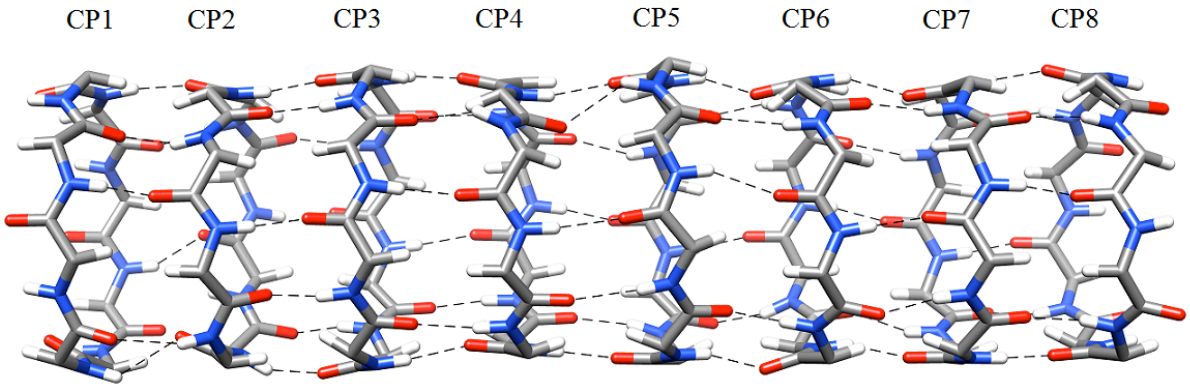

Figure 1: Schematic illustration of the CPNT with sequential naming of monomeric CP units. The intermolecular H-bonding interactions are shown as dotted lines. Side chains are not shown for clear visibility.

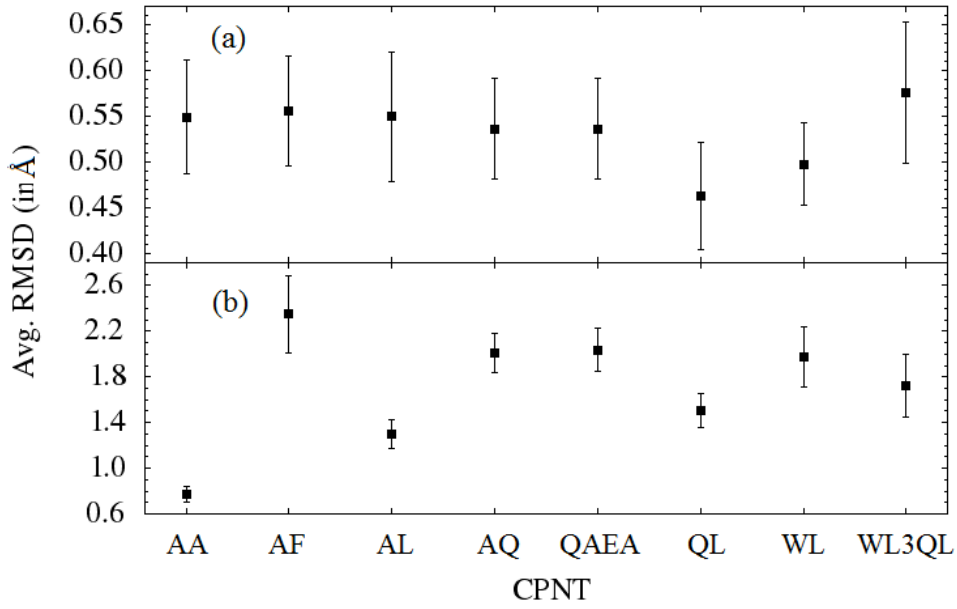

Figure 2: Average RMSD values of backbone (a) and side chain (b) heavy atoms of different model systems with standard deviation.

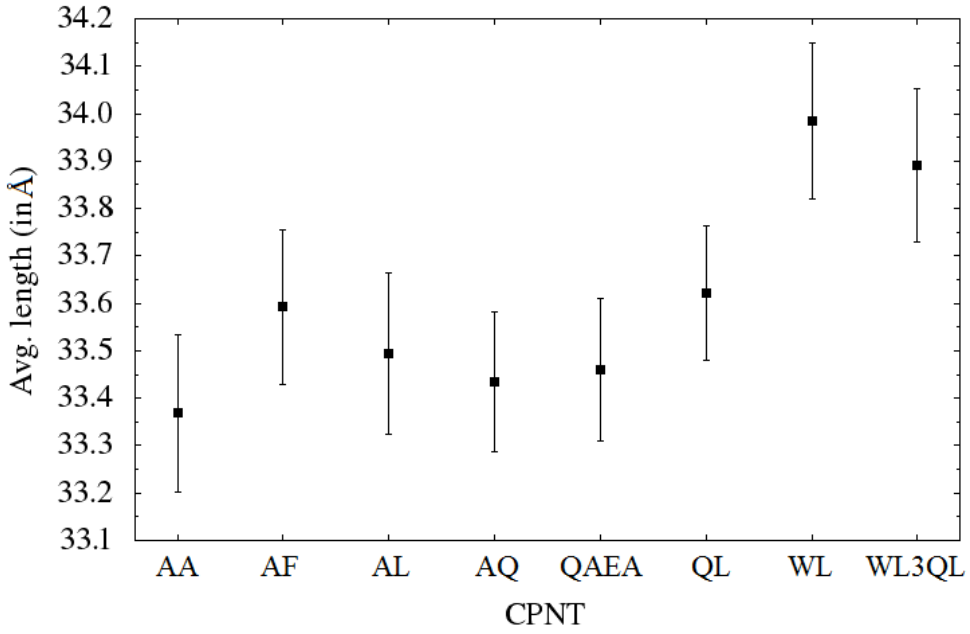

Figure 3: Average length with respect to the centre of mass of the Cα plane distance between the first and last CP unit of the CPNTs considered.

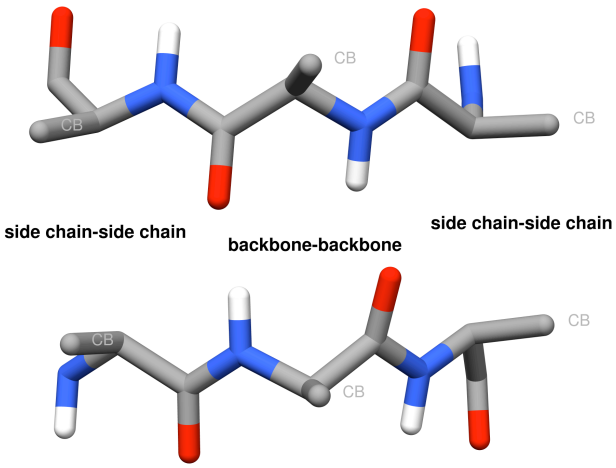

Figure 4: Truncated CP dimer showing the intermolecular interactions (side chain hydrogen atoms are not shown for clear visibility).

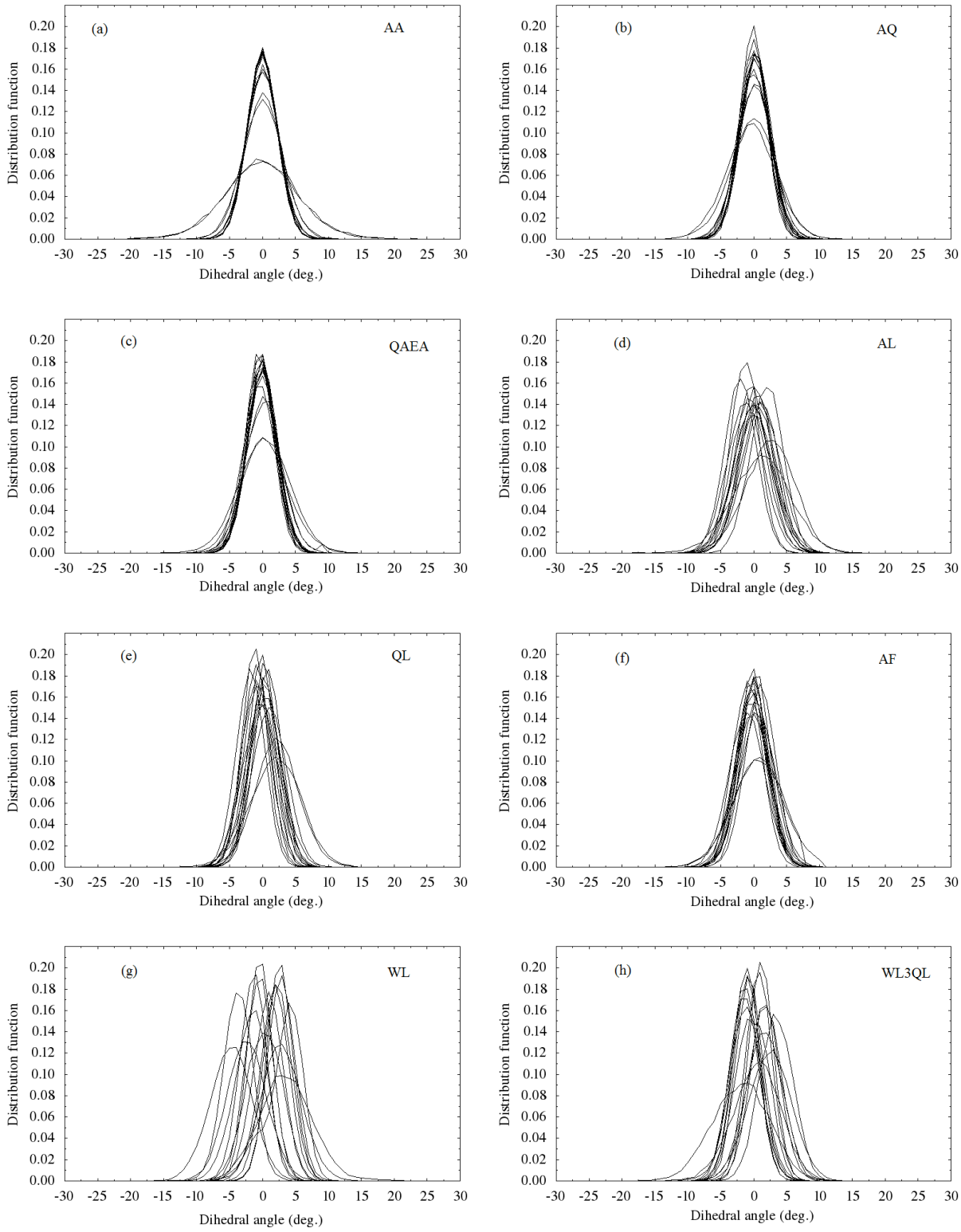

Figure 5: normalized distribution of alternate Cα atoms dihedral angles of each CP unit in the different models systems.

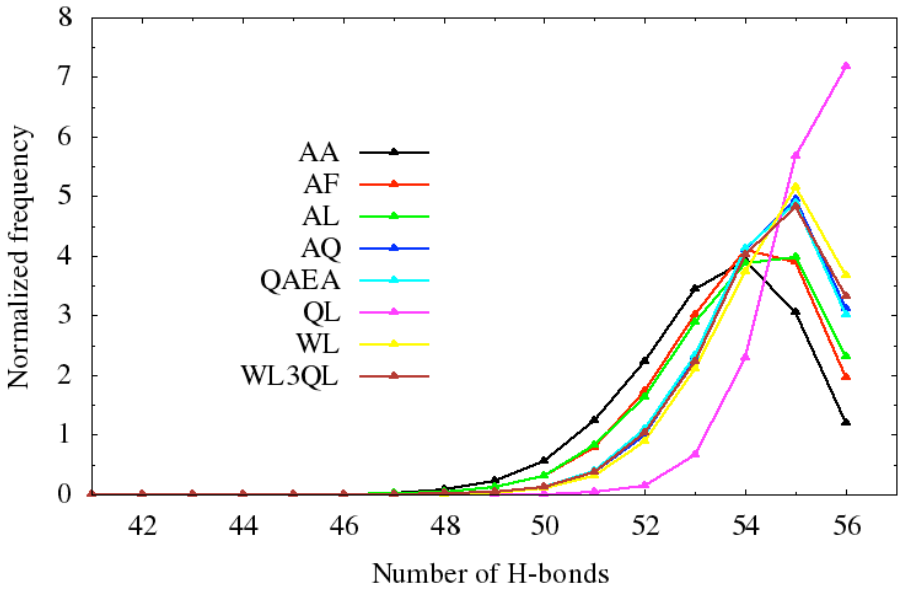

Figure 6: Normalized frequency of backbone-backbone H-bonds observed between different CP units of all model systems.

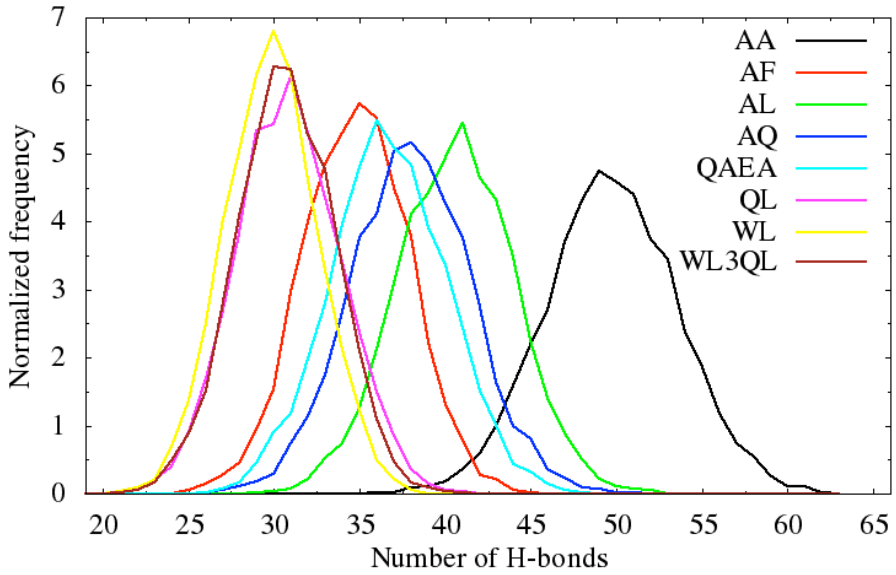

Figure 7: Normalized frequency of H-bonds observed between the backbone carbonyl O atom and water molecules of all model systems.

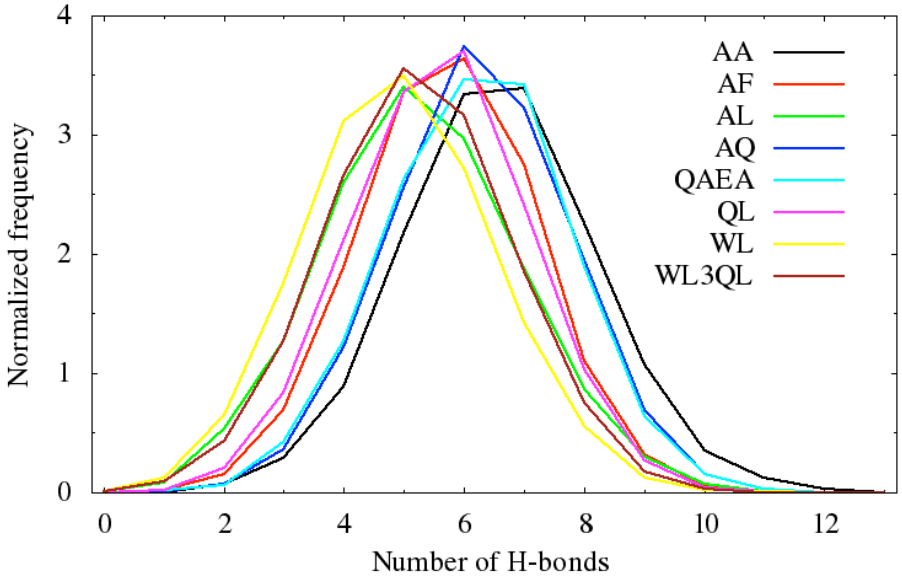

Figure 8: Normalized frequency of H-bonds observed between the backbone amide groups and water molecules of all model systems.

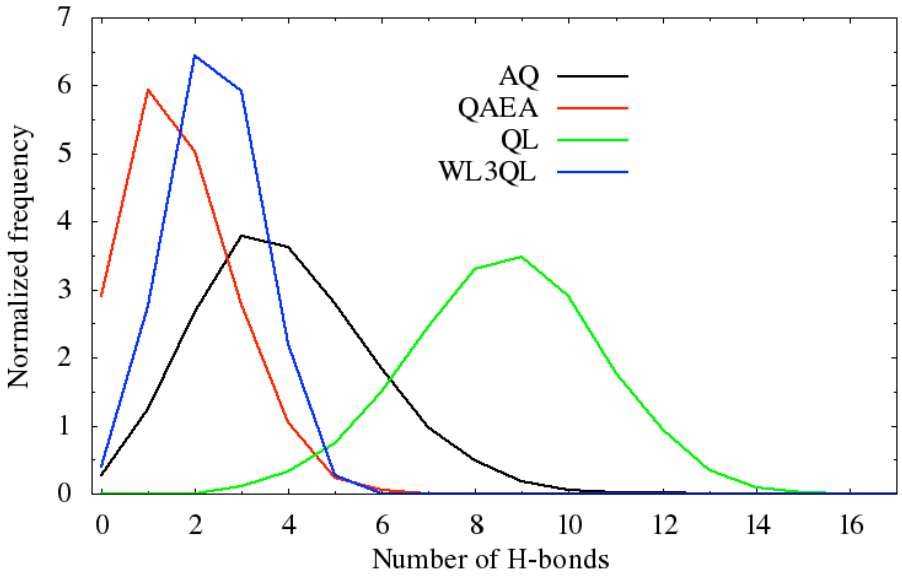

Figure 9: Normalized frequency of H-bonds between the side chains of the Gln residues in applicable CPNTs.

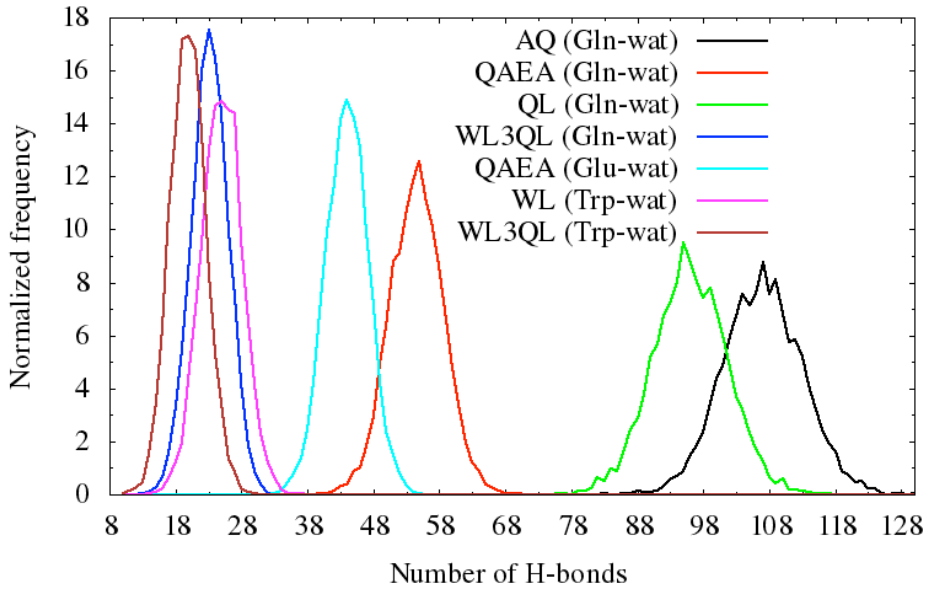

Figure 10: Normalized frequency of H-bonds between the side chains and water molecules in applicable CPNTs.